

Aufgabe 1:

Im Datensatz `leprosylong.txt` sind die Anzahl von Lepra-Bazillen vor und nach der Behandlung mit Antibiotika enthalten. Es existieren für jeden Patienten folgende Variablen: Medikament (`Drug`), Anzahl der Bazillen (`count`) und Zeit (`time`). Medikament C entspricht einer Placebo-Behandlung. `time 0` entspricht dem Zustand vor der Behandlung, `time 1` dem Zustand nach der Behandlung.

Benutzen Sie den folgenden Code um den Datensatz einzulesen:

```
R>url <- "http://www.statistik.lmu.de/institut/lehrstuhl/semwiso/longitudinal_ss08/"
R>leprosy <- read.table(paste(url, "download/leprosylong.txt",
+   sep = ""), header = T)
R>leprosy$Drug <- factor(leprosy$Drug, levels = c("C", "A", "B"))
```

- (a) Bestimmen Sie Mittelwert und Varianz der Anzahlen der Bazillen vor und nach der Behandlung für jede Medikamentengruppe und plotten Sie die Verläufe getrennt nach Gruppen.
- (b) Passen Sie mit Hilfe verallgemeinerter Schätzgleichungen (GEE) ein Modell an die Daten an, welches als Einflussgrößen die Variablen Zeit und eine Interaktion von Behandlung mit Zeit hat. Gehen Sie von unabhängigen Beobachtungen aus. Was verbirgt sich hinter den 'robusten' und den 'naiven' Schätzungen der Varianzen der Parameter? Beurteilen Sie wie sinnvoll die getroffene Annahme für die Korrelationsstruktur der Daten ist.
Hinweis: Benutzen Sie das `gee`-Paket.
- (c) Schätzen Sie das Modell erneut unter der Annahme dass Beobachtungen am selben Patienten nicht unabhängig sind.
- (d) Überprüfen Sie mit Hilfe der (multivariaten) Wald-Statistik, ob Behandlungseffekte der beiden Antibiotika vorliegen (Nullhypothese: $\beta_3 = \beta_4 = 0$). Interpretieren Sie die Parameterschätzer.
- (e) Überprüfen Sie, ob die Behandlungseffekte der beiden Antibiotika gleich sind (Nullhypothese: $\beta_3 = \beta_4$).
- (f) Fassen Sie beide Antibiotikabehandlungen zusammen und passen Sie das Modell entsprechend an. Geben sie ein Konfidenzintervall für den Behandlungseffekt an.